Masses and Ages for 230,000 LAMOST Giants, via Their Carbon and Nitrogen Abundances

Anna Y. Q. Ho[1,2], Hans-Walter Rix[2], Melissa K. Ness[2], David W. Hogg[2,3,4,5], Chao Liu[6], Yuan-Sen Ting (丁源森)[7]

ah@astro.caltech.edu

ABSTRACT

We measure carbon and nitrogen abundances to $\lesssim 0.1$ dex for 450,000 giant stars from their low-resolution ($R \sim 1800$) LAMOST DR2 survey spectra. We use these [C/M] and [N/M] measurements, together with empirical relations based on the APOKASC sample, to infer stellar masses and implied ages for 230,000 of these objects to 0.08 dex and 0.2 dex respectively. We use *The Cannon*, a data-driven approach to spectral modeling, to construct a predictive model for LAMOST spectra. Our reference set comprises 8125 stars observed in common between the APOGEE and LAMOST surveys, taking seven APOGEE DR12 labels (parameters) as ground truth: $T_{\rm eff}$, $\log g$, [M/H], [α/M], [C/M], [N/M], and A_k. We add seven colors to the Cannon model, based on the *g, r, i, J, H, K, W1, W2* magnitudes from APASS, 2MASS & WISE, which improves our constraints on $T_{\rm eff}$ and $\log g$ by up to 20% and on A_k by up to 70%. Cross-validation of the model demonstrates that, for high-S/N objects, our inferred labels agree with the APOGEE values to within 50 K in temperature, 0.04 magnitudes in A_k, and < 0.1 dex in $\log g$, [M/H], [C/M], [N/M], and [α/M]. We apply the model to 450,000 giants in LAMOST DR2 that have *not* been observed by APOGEE.

[1]Cahill Center for Astrophysics, California Institute of Technology, MC 249-17, 1200 E California Blvd, Pasadena, CA, 91125, USA

[2]Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

[3]Simons Center for Data Analysis, 160 Fifth Avenue, 7th floor, New York, NY 10010, USA

[4]Center for Cosmology and Particle Physics, Department of Phyics, New York University, 4 Washington Pl., room 424, New York, NY, 10003, USA

[5]Center for Data Science, New York University, 726 Broadway, 7th floor, New York, NY 10003, USA

[6]Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Datun Road 20A, Beijing 100012, China

[7]Harvard Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

This demonstrates that precise individual abundances can be measured from low-resolution spectra, and represents the largest catalog of [C/M], [N/M], masses and ages to date. As as result, we greatly increase the number and sky coverage of stars with mass and age estimates.

Subject headings: methods: data analysis — methods: statistical — stars: abundances — stars: fundamental parameters — surveys — techniques: spectroscopic

1. Introduction

An empirical description of the Milky Way's present structure and formation history requires accurate and consistent age estimates for large samples of stars distributed throughout the galaxy. Although we have recently entered an era of extensive spatial, kinematic, and chemical information beyond the solar neighborhood, comparably extensive age constraints remain elusive.

Stellar age is a property that must be inferred from observations with the help of stellar evolution models; generally, it cannot be measured "directly". Therefore, results are inherently limited by the applicability and accuracy of the model used (see Soderblom (2010) for a comprehensive review.) As stellar ages are difficult to measure directly, abundances such as [Fe/H] and [α/Fe] are commonly used as an age-dating proxy (e.g. via making maps of mono-age populations; see Rix & Bovy (2013) and Bovy et al. (2015)) because the determination of photospheric abundances from spectra is more straightforward.

Unfortunately for Milky Way studies, the population of stars that is most readily observable throughout the galaxy — red giant stars — is also the one for which it is particularly challenging to estimate ages. The standard technique of isochrone fitting is impractical in this regime, because there is too much uncertainty both in stellar parameter measurements and in the model isochrones. In other words, stars with very different ages can have very similar atmospheric parameters and luminosities (Rix & Bovy 2013; Soderblom 2010).

Instead, the key to age-dating red giant stars is mass. Because the red giant phase is so short, the age of a star is essentially its main sequence lifetime, which is set by the star's mass and metallicity (Soderblom 2010). Given mass and metallicity, one can estimate age using isochrones, e.g. by interpolating between them.

Recently, asteroseismology has made it possible to measure masses for giants out to large distances. The *Kepler* mission (Koch et al. 2010; Gilliland et al. 2011) has conducted long-cadence photometry for over ten thousand giants along a pencil-beam through the Galaxy

(Stello et al. 2013). From detailed light curves one can measure two characteristic variability frequencies that directly probe the (age-dependent) structure of the stellar interior: $\nu_{\max}$ is the frequency corresponding to the maximum oscillation power, and $\Delta\,\nu$ is the frequency spacing between two consecutive modes of the same spherical degree. This approach is especially effective for giants because they have higher densities and thus a larger sound speed, which makes these (acoustic) oscillations more pronounced (Soderblom 2010). Together with a measurement of the star's $T_{\rm eff}$, and the solar values $\nu_{\max,\odot}$ and $\Delta\,\nu_{\odot}$, the mass of the star can be estimated using seismic scaling relations. Note that these scaling relations are based on Sun-like stars, and may not be suitable for metal-poor stars (Epstein et al. 2014).

Furthermore, the population of stars with asteroseismic measurements is spatially limited. Ness et al. (2016) and Martig et al. (2016) greatly expanded the spatial coverage of giants with age estimates by determining masses spectroscopically: they showed that the masses (and implied ages) of post dredge-up giants can be measured from high-resolution infrared (APOGEE, $R \approx 22{,}500$) spectra, and determined a model of mass and age as a function of $T_{\rm eff}$, $\log g$, [M/H], [C/M], and [N/M] values (see Tables A2 and A3 in Martig et al. (2016)). Their work increased the sample of giant stars with known ages to 70,000, the largest (and most spatially extended) sample of stellar ages to date.

In this work, we set out to extend this spectroscopic mass & age work to LAMOST, the largest ongoing stellar spectroscopic survey. LAMOST represents a large expansion over APOGEE in area coverage (LAMOST stars are measured away from the disk, unlike APOGEE), sample size, and parameter range (in particular, [Fe/H]). Ho et al. (2016) have shown that basic parameters ($T_{\rm eff}$, $\log g$, [Fe/H], and [α/M]) consistent with APOGEE values can be determined directly from LAMOST spectra, using *The Cannon*.

The Cannon (Ness et al. 2015) is a data-driven method for measuring stellar "labels" (a term that refers collectively to all attributes of a star, e.g. physical parameters and element abundances) from stellar spectra in the context of large spectroscopic surveys. It has shown promise for bringing qualitatively different stellar surveys onto a consistent physical scale, and for transferring label systems from one survey to another. In Ho et al. (2016), for example, *The Cannon* was used to transfer labels from a high-resolution, high-S/N survey (APOGEE) to a low-resolution, modest-S/N survey (LAMOST), enabling the measurement of the first-ever [α/M] values from LAMOST spectra, and the largest and most spatially-extended sample of [α/M] values to date ($\sim$450,000 giants). Of course, this data-driven approach can only be applied to a subset of the LAMOST dataset, restricted by the overlap in label space with APOGEE: because the set of objects in common between APOGEE and LAMOST is entirely giants, our model is limited in its applicability to giants, which is only 20% of the LAMOST DR2 dataset.

The Cannon operates in two s.png: a training step and a test step. For a complete description of the methodology, see Ness et al. (2015); we summarize briefly here. In the training step, *The Cannon* uses a reference set of objects observed in common between the surveys of interest to construct a predictive model of spectra independently at each wavelength. For example, for a set of objects measured in common between Survey A and Survey B, *The Cannon* might fit a model that predicts every pixel in a Survey A spectrum given Survey B labels. In the test step, this model can be used to infer *new* labels directly from Survey A spectra that are by construction on the Survey B label scale. *The Cannon* uses no explicit physical stellar models, is very fast, and achieves comparable accuracy to existing survey pipelines using significantly lower SNR spectra; it requires only a set of objects observed in common between the surveys.

Taken together, the work in Martig et al. (2016), Ness et al. (2016), and Ho et al. (2016) raises the prospect that ages could be determined for a large sample of LAMOST giants. In theory, it seems plausible that mass (and implied age) information could be encoded in optical spectra. After all, in the near-IR, mass is encoded in CN and CO molecular regions; as [C/N] and [C/H] features are prominent in the blue parts (~ 4100Å) of giant spectra (e.g. Martell et al. (2008)) it seems plausible that this information could be encoded in LAMOST spectra too. Recent theoretical work by Salaris et al. (2015) and Martig et al. (2016) lends physical justification to why these features should be indicative of mass: during a star's main sequence lifetime, the CNO cycle in its core determines the final relative abundances of carbon and nitrogen. After arriving on the giant branch, the material in the core is dredged up to the surface via convective mixing. The depth of the convective envelope, and the [C/N] ratio in the core, is determined by the mass of the star. Thus, in giants that have undergone dredge-up once (that is, they have not undergone additional convective mixing) the [C/N] ratio observed in the photosphere is (together with metallicity) highly predictive of mass.

However, [C/M] and [N/M] have not previously been measured from low-resolution resolution ($R \lesssim 5000$) spectra. In this work, we extend the APOGEE-LAMOST label transfer work of Ho et al. (2016) by two additional labels ([C/M] and [N/M]) to learn about the information content of LAMOST spectra. We use the theoretical relations in Martig et al. (2016) to determine masses and ages for as many giant stars as possible, restricted primarily by the parameter regime in which the relations are applicable. This will enable us to measure the largest sample of stellar ages to date.

2. Label Transfer Using *The Cannon*

Our procedure for transferring labels from APOGEE to LAMOST using *The Cannon* closely resembles the work of Ho et al. (2016). Here, too, the data consist of spectra from LAMOST, which we again normalize in a consistent way using a running Gaussian, and labels from APOGEE DR12. As before, the spectral model is quadratic in the labels. There are, however, important differences in (and new components to) our modeling: in the labels that we use, in the reference objects that we use to train the model, in spectral regions that we mask out, and in the incorporation of photometry.

Here, our model is quadratic in seven labels instead of the original five labels: we use $T_{\rm eff}$, $\log g$, [M/H], [C/M], [N/M], [α/M], and K-band extinction $A_{\rm k}$. Because we will eventually use the relations in Martig et al. (2016) to translate our carbon and nitrogen abundances into age estimates, our labels need to be on the same scale as those that were used to fit for the relations. Thus, whereas we used the calibrated DR12 parameters in Ho et al. (2016) (those in the `PARAM` array) in this case we use the raw, uncalibrated values from the `FPARAM` array.

Furthermore, we do not use the full reference set of 9956 objects from Ho et al. (2016), because some of these have unreliable [C/M] and [N/M] reference labels. Following Martig et al. (2016), we excise objects that have both $T_{\rm eff} > 4550$ and $-1 < [{\rm M/H}] < -0.5$ (743 objects) in order to eliminate objects with only an upper limit measurement on [C/M] and a lower limit on [N/M]. In addition, Martig et al. (2016) found that the minimum [C/M] possible to measure is on the level of -0.4 to -0.5 dex, so we also exclude objects with $[{\rm C/M}] < -0.4$ (40 objects). This left us with 9173 out of the original 9956.

In Ho et al. (2016), we fit an independent spectral model at *every* spectral pixel. However, there are features in LAMOST spectra that arise from effects in a different velocity system from that of the star: for example, Diffuse Interstellar Bands (DIBs) and the Na I doublet are interstellar absorption features originating from intervening material. We noticed by examining the leading coefficients of an initial Cannon model (see Figure 3) that *The Cannon* was "learning" to use these features to predict labels intrinsic to the star, particularly [α/M], and that this introduced radial velocity-dependent systematic errors into the label estimates[1]. The leading coefficients of an initial *The Cannon* model also indicated that the imperfectly corrected telluric bands, originating in the Earth's atmosphere, left small, but significant, velocity-dependent effects in the rest-frame spectra of the stars.

---

[1]As α-enhanced stars have a different line-of-sight velocity distribution, velocity and [α/M] may well be correlated

To prevent *The Cannon* from using these features spuriously, we masked them out by setting the inverse variances corresponding to these pixels to be zero. To be conservative, roughly half of each spectrum was masked, but because the most important features for our labels of interest were preserved, this still improved our label estimates.

Masking out the interstellar absorption features took out most of the spectral information on A_k, which originated in the DIB strength. But of course, many-band photometry of a star encodes a combination of its effective temperature and its reddening. We found that incorporating photometry not only enabled us to accurately and precisely predict A_k for our reference objects, but also to improve the precision of our estimate of T_{eff}, particularly for lower-S/N spectra. It also improved the precision of our estimate of $\log g$, presumably because measurements of T_{eff} and $\log g$ are highly covariant for spectra of this resolution due to blending (see e.g. Ting et al. 2016b; submitted to ApJ). The scatter in A_k decreased by 70% in the lowest-S/N spectra, by 50% in spectra with $50 < S/N < 100$, and by 25% in spectra with $S/N > 100$. For low-S/N spectra, the scatter in T_{eff} and $\log g$ improved by up to 20%, although for $S/N \gtrsim 50$ the difference was negligible.

We incorporated photometry as follows: we took magnitudes (and the associated uncertainties) from eight bands, taken from APASS DR9 (Henden & Munari 2014; Henden et al. 2016), 2MASS (Cutri et al. 2003; Skrutskie et al. 2006) and WISE (Wright et al. 2010): *g, r, i, J, H, K, W1* (3.4 μm), *W2* (4.6 μm). From these, we constructed seven colors: *g-r*, *r-i*, *i-J*, *J-H*, *H-K*, *K-W1*, and *W2-W1*. For each reference object, we added its seven colors as "pixels" to its spectrum: the color as the "flux" and the uncertainty as the "error bar." Using colors restricted us to the set of reference objects with APASS, 2MASS & WISE magnitudes: 8472 of the 9173.

We emphasize that *The Cannon* builds a model of spectra but is agnostic to whether the value of a "spectral pixel" is a true flux value or simply another observed property of the star that is sensitive to the labels of interest. Note, however, that our quadratic model is probably not sufficiently complex for the photometric data; we would expect improvements for a model that is more complex at the photometric pixels relative to the spectroscopic pixels.

2.1. Results from Cross-Validation

As in Ho et al. (2016), we evaluate the accuracy and precision of our model using a "leave-$\frac{1}{8}$-out" cross-validation test. We split the 8472 reference objects into eight groups, by assigning each one a random integer between 0 and 7. We leave out each group in turn, and train a model on the remaining seven groups. We then apply that model to infer new labels for the group that was left out. We use the results of the cross-validation to determine which objects are appropriate reference objects for training the model: we excise objects whose difference from the training value is greater than four times the scatter in that label, leaving 8125 objects. We train the model on these 8125 objects and re-run the cross-validation. We also use the model to infer labels for the 347 objects excised from the training, in order to properly account for all of the objects in the following error analysis.

At the end of this process, each of the 8472 objects has a new set of labels determined by *The Cannon*, from a model that was *not* trained using that object. Figure 1 shows the comparison of these Cannon-inferred "test" labels with the reference labels used in training, for high S/N objects; there is a significant decrease in scatter compared to the corresponding figure in Ho et al. (2016) (Fig. 6). For objects with S/N > 100, the labels are consistent with the APOGEE training values to within 53 K in $T_{\rm eff}$, 0.11 dex in $\log g$, 0.05 dex in [M/H], 0.06 dex in [C/M], 0.09 dex in [N/M], 0.03 dex in [α/M], and 0.04 mag in $A_{\rm k}$. These are comparable to, or within, the stated ASPCAP uncertainties (Holtzman et al. 2015).

Figure 2 shows the scatter in different bins of S/N (where S/N is the median value of formal S/N across all pixels in the spectrum) in all of the labels except for $A_{\rm k}$ (which is primarily determined from the additional photometric pixels, not taken into account in determining the S/N). By construction, as a result of this data-driven label transfer, there is significant improvement in agreement with the APOGEE values over those from the LAMOST pipeline.

2.2. Astrophysical Verification of the Spectral Model

A key strength of *The Cannon* is the physical interpretability of the spectral model. An independent model is fit at every pixel of the spectrum, so each pixel has a set of model coefficients as well as a model scatter term. The leading (linear) coefficient in each label can be thought of as a proxy for how sensitive a particular spectral pixel is to that particular label; thus, each label has a wavelength-dependent indicator of which spectral regions are most informative. Each pixel also has a model scatter term; this is the intrinsic variance in the model at that pixel, as distinct from the observational variance. In other words, it is the expected deviation from the model at that particular pixel, in the limit of a perfect measurement.

These are shown in Figure 3, in the part of the spectrum found by *The Cannon* to be most predictive of labels (the blue end, $\sim 4000 - 5800$ Å), together with the scatter in the model. These linear coefficients are the first derivative of the model at a set of fiducial stellar

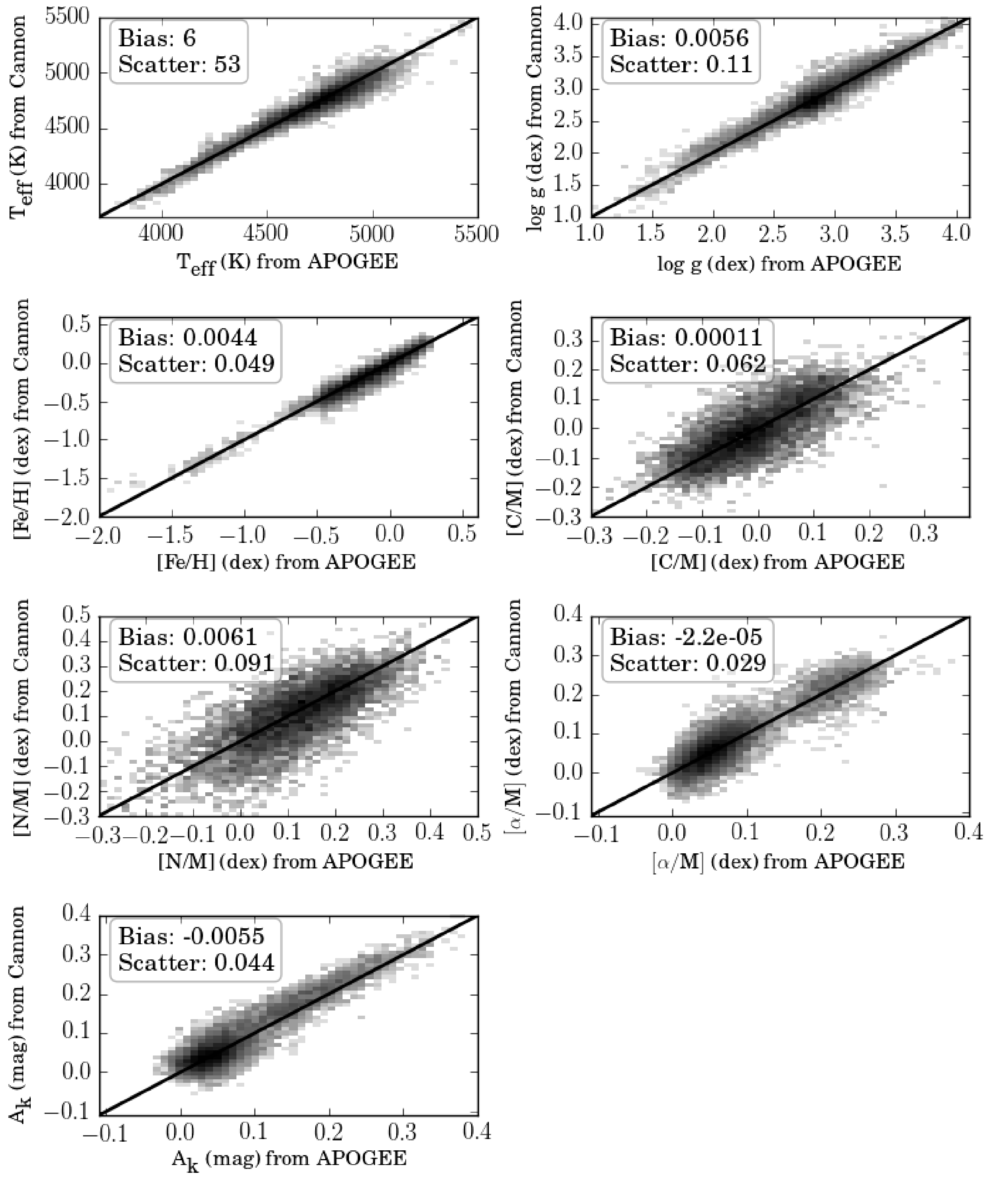

Fig. 1.— Results from cross-validation of *The Cannon*'s label transfer from APOGEE to LAMOST, for spectra with S/N > 100. Shown are the APOGEE labels used in training the model, compared to the labels inferred by *The Cannon* in the test step. The improvement in precision over the results in Ho et al. (2016) reflects changes we made to the model described in 2. The low bias and scatter in [C/M] and [N/M] demonstrates that these abundances can in fact be measured from low resolution LAMOST spectra.

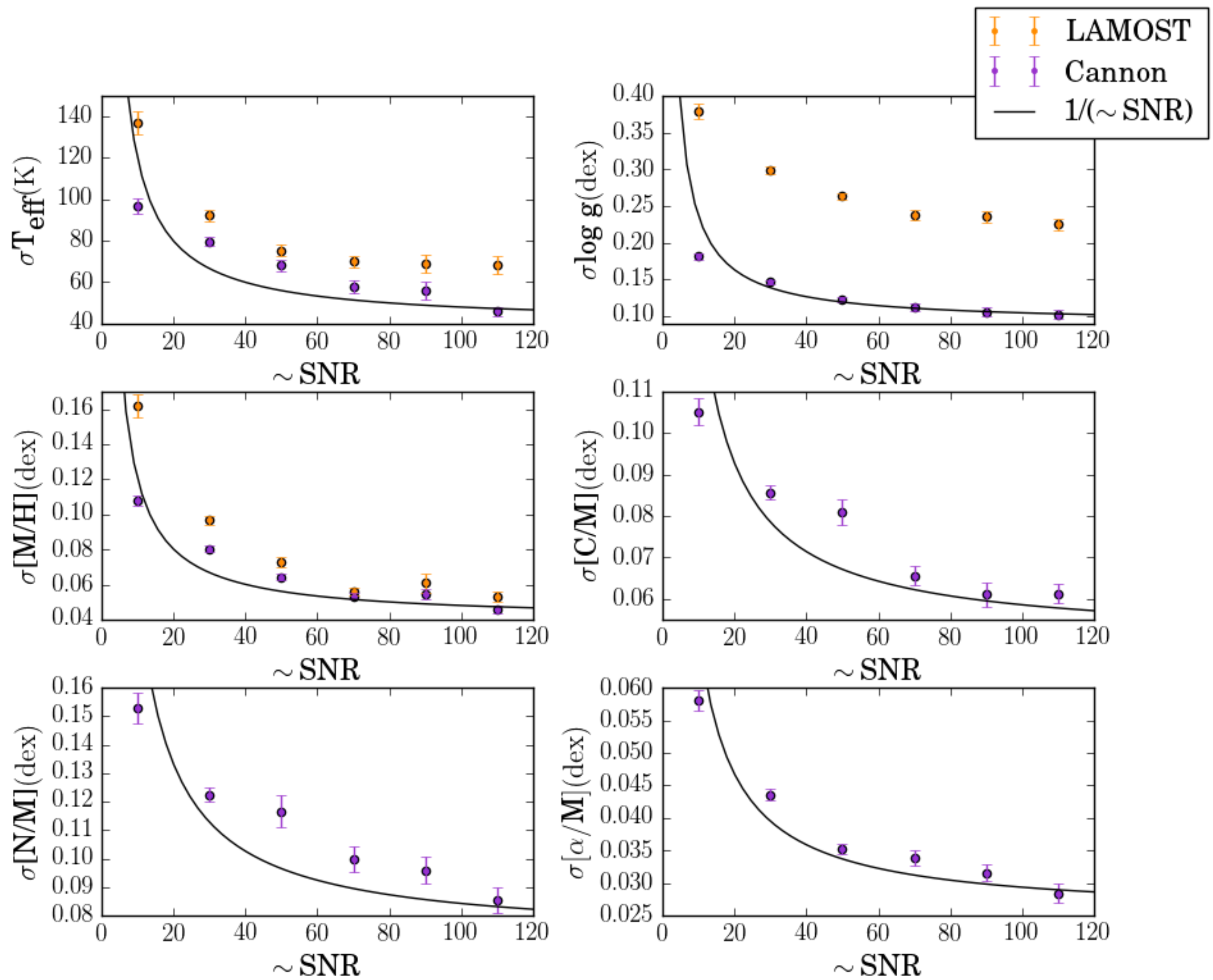

Fig. 2.— The S/N-dependence of the scatter between APOGEE DR12 labels and the corresponding labels measured from LAMOST spectra by *The Cannon* (purple points) and the LAMOST pipeline (yellow points), for 8472 objects. By construction, the labels measured by *The Cannon* are in closer agreement with the APOGEE values for $\mathrm{T_{eff}}$, $\log \mathrm{g}$, and [M/H], and the model behaves well with decreasing S/N. Note that we are using our own formal measurement of $\sim$ S/N, not the reported LAMOST error bar.

parameters; in this case, we pivot the model around the mean value in each label across the training set: $\mathrm{T_{eff}} = 4687\,\mathrm{K}$, $\log \mathrm{g} = 2.83\,\mathrm{dex}$, $[\mathrm{M/H}] = -0.19\,\mathrm{dex}$, $[\alpha/\mathrm{M}] = -0.01\,\mathrm{dex}$, $[\mathrm{C/M}] = 0.10\,\mathrm{dex}$, $[\mathrm{N/M}] = 0.09\,\mathrm{dex}$, $\mathrm{A_k} = 0.09\,\mathrm{mag}$. To facilitate comparison, each derivative has been scaled by the approximate error in the corresponding label: $\delta\mathrm{T_{eff}} \sim 91.5\,\mathrm{K}$, $\delta\log \mathrm{g} \sim 0.11\,\mathrm{dex}$, $\delta[\mathrm{M/H}] \sim 0.05\,\mathrm{dex}$, $\delta[\alpha/\mathrm{M}] \sim 0.05\,\mathrm{dex}$, $\delta[\mathrm{C/M}] \sim 0.04\,\mathrm{dex}$, $\delta[\mathrm{N/M}] \sim 0.07\,\mathrm{dex}$ (Holtzman et al. 2015).

The new labels in this work are [C/M] and [N/M], so we should demonstrate that *The Cannon* measures these values from physically plausible spectral signatures. Ting et al. 2016b (submitted to ApJ) quantified how the precision with which various abundances (including [C/H] and [N/H]) can be measured varies as a function of survey resolution. They showed that low-resolution ($R < 10,000$) spectra, such as those from LAMOST, have the same theoretically achievable uncertainties per stellar label as medium-resolution ($10,000 < R < 50,000$) spectra, under the following assumptions: equal exposure time (so that a low-resolution spectrum has higher S/N per resolution element), an equal number of detector pixels (so that low-resolution spectra have more extensive wavelength coverage), and a constant sampling per resolution element. These predictions are based in part on gradient spectra calculated using Kurucz models (Kurucz 1970, 1993, 2005; Kurucz & Avrett 1981), which are assumed to be perfect. Of course, this aspect does not pertain to data-driven models (see the discussion in Ting et al. (2016a)).

To make a direct comparison between the Cannon model and theoretical predictions, we calculate gradient spectra and compare them to the gradient spectra calculated from Kurucz models by Ting et al. (2016a). Gradient spectra are a quantification of how much the flux at a given wavelength changes with changes to a given label: in other words, it characterizes the sensitivity or information content of each wavelength for a given label. Following Equation 2 in Ting et al. 2016b, gradient spectra are calculated as follows:

$$\nabla_{\ell} f_{\mathrm{model}}(\lambda, \ell_i) = \frac{f_{\mathrm{model}}(\lambda, \ell_i + \Delta\ell_i) - f_{\mathrm{model}}(\lambda, \ell_i)}{\Delta\ell_i} \tag{1}$$

where $f_{\mathrm{model}}(\lambda, \ell_i)$ represents a model spectrum across wavelengths λ, generated using a set of labels ℓ_i. A fractional change in a particular label $\Delta\ell_i$ results in a fractional change in the spectrum $\nabla_{\ell} f_{\mathrm{model}}$ at each wavelength λ. In other words, to study sensitivity of a spectral region to a particular label, one changes the value in that label and calculates the fractional change in the flux in that region.

We have two sets of model spectra: one from the Cannon model as described in Section 2, and one from the Kurucz models, in both cases generated using labels for a solar metallicity K-giant ($\mathrm{T_{eff}}$=4800, log g=3.5). We use each of these models to calculate a gradient spectrum

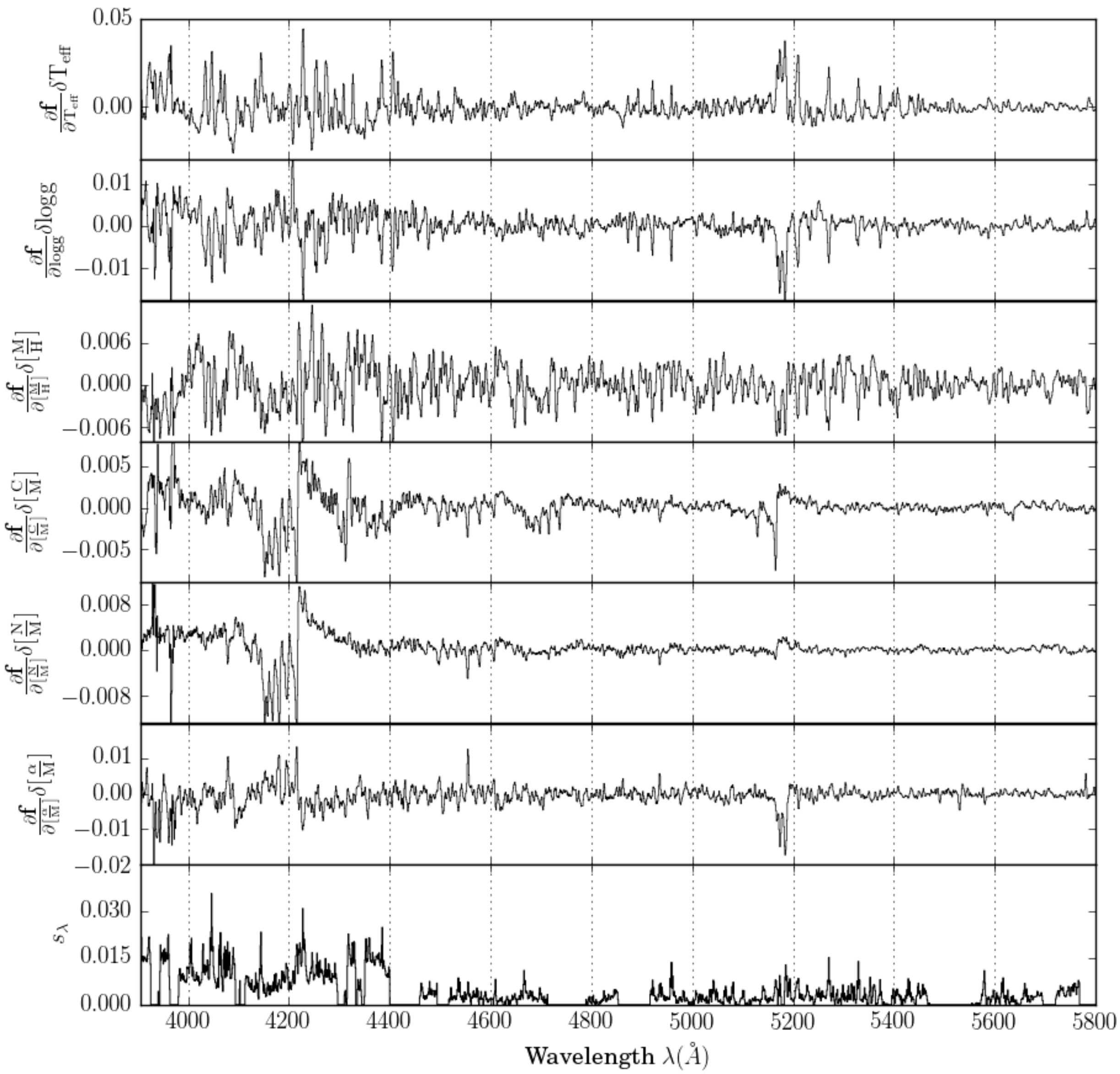

Fig. 3.— Leading (linear) coefficients and scatter from the best-fit spectral model, determined by *The Cannon* using the 8125 reference objects. The leading coefficients can be thought of as a proxy for how sensitive each pixel in the spectrum is to each of the labels; to facilitate comparison, each has been scaled by the approximate error in that label (Holtzman et al. 2015). The scatter term s_λ is the variance intrinsic to the model; it quantifies each pixel's expected deviations from the model in the limit of a perfect measurement. We display the model coefficients and scatter blueward of 5800 Å because this was found by *The Cannon* to be the region with the most sensitive features.

for [C/M] by varying [C/M] by 0.2 dex, and a gradient spectrum for [N/M] by varying [N/M] by 0.2 dex. For a more direct comparison, we normalize the theoretical gradient spectra the same way as the LAMOST spectra.

Figure 4 shows the Cannon model gradient spectra overlaid with the Kurucz model gradient spectra from Ting et al. 2016b, for the CN and CH molecular features in the wavelength range 4100-4400 Å (see Martell et al. (2008)). The panel on the left shows the gradient spectra for carbon, and the panel on the right shows the gradient spectra for nitrogen. As the theoretical gradient spectra (red) were generated purely from physical models, and the Cannon gradient spectra (black) represent an entirely data-driven relationship between these wavelength regions and abundances from APOGEE, the qualitative similarity between them is gratifying. There are clearly some quantitative differences, but we simply seek to demonstrate here that the signatures of carbon and nitrogen from the data-driven Cannon model come from astrophysically sensible wavelength regions, such as the 4215 Å CN band. Furthermore, the differences between the two panels demonstrates that we are measuring each element from distinct features, not simply correlations between the two (e.g. the carbon-sensitive CH (G) band, not present in the nitrogen signature). The fact that [C/M] and [N/M] share regions of sensitivity does not mean that they are degenerate; they may be covariant, but can still be independently measured when fit for simultaneously (see Ting et al. 2016b, submitted).

3. From [C/M] and [N/M] to Mass and Age

To transform [C/M] and [N/M] to mass and age, we use the formulas characterized by the coefficients in Table A2 and Table A3 of Martig et al. (2016), which are in turn based on asteroseismic mass measurements for stars with [C/M] and [N/M] measurements. These relations are only applicable within a certain range of label values, restricting the number of objects for which we can infer masses and ages via their [C/M] and [N/M]. Although we infer [C/M] and [N/M] for the full set of 454,180 test objects described in Ho et al. (2016), we apply the following cuts (following Martig et al. (2016)) which leave 230,901 objects suitable for applying the formula. This sets the primary restriction on the size of our mass and age catalog; for example, although the LAMOST data contain a large population of low-metallicity outer disk stars, we cannot estimate masses and ages for those objects.

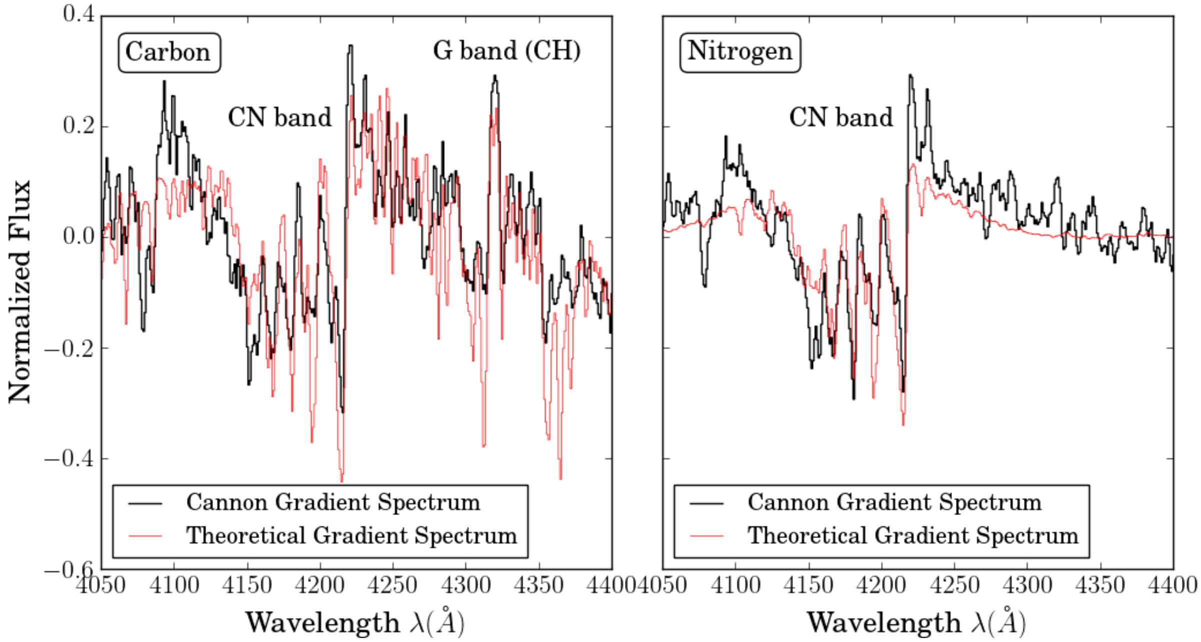

Fig. 4.— Gradient spectra for carbon (left panel) and nitrogen (right panel) calculated using two different models: theoretical Kurucz models (red line) and the Cannon model (black line). All model spectra were generated using K-giant (T_{eff}=4750, $\log g$=2.5) and solar metallicity values, stepping over 0.2 dex in [C/M] and [N/M]. For a better comparison, the Kurucz gradient spectra were normalized the same way as the LAMOST spectra, as described in Section 2. The qualitative similarity between the black and red lines demonstrates that the Cannon measurements of [C/M] and [N/M] are coming from astrophysically sensible spectral regions, e.g. the 4215 Å CN band. Furthermore, the difference between the left and right panels demonstrates that [C/M] and [N/M] are being measured independently, not just via correlations: for example, the nitrogen signature does not include any changes in the G (CH) band.

$$\begin{cases} -0.8 < [\mathrm{M/H}] < 0.25 \\ 4000 < \mathrm{T_{eff}} < 5000 \\ 1.8 < \log \mathrm{g} < 3.3 \\ -0.25 < [\mathrm{C/M}] < 0.15 \\ -0.1 < [\mathrm{N/M}] < 0.45 \\ -0.05 < [\alpha/\mathrm{M}] < 0.3 \\ -0.1 < [(\mathrm{C+N})/\mathrm{M}] < 0.15 \\ -0.6 < [\mathrm{C/N}] < 0.2 \end{cases}$$

The estimated masses and ages have uncertainties that come from the intrinsic scatter of the relation (Martig et al. 2016) and the individual stellar label uncertainties. To estimate the latter, we sample from each star's label *pdf* 100 times, approximating each distribution as a Gaussian with a standard deviation equal to the scatter in that label. That scatter is a function of signal to noise (see Figure 2), so for each object we take the standard deviation of its Gaussian spread to be the scatter associated with that signal to noise bin. Thus, each object has a distribution in mass and age. We take the mass and age value to be the median of that distribution, and estimate the uncertainty using the half-width of the central 68^{th} percentile.

This procedure does not account for the errors in the training labels nor the scatter in the Martig et al. (2016) relation. There are, however, additional systematic errors from the age relation in Martig et al. (2016) that are not taken into account here. Note that these are all distinct from, an in addition to, the formal error from the Cannon model fit.

We provide a catalog of all our inferred labels, including mass and age; an excerpt is shown in Table 1. We provide [C/M] and [N/M] for all 454,364 objects, but mass and age only for the 230,901 of those that fall within the label space of Martig et al. (2016). We also provide flags indicating whether an object was used as a reference object and whether the object falls within the label range of Martig et al. (2016). We also provide the formal S/N for the spectrum of each object and the reduced chi squared of the fit, which is the chi squared divided by the approximate number of pixels in each spectrum ($\sim$ 1800). Note that the S/N and the reduced chi squared are both low by roughly a factor of three; see the discussion in Section 4.1 of Ho et al. (2016). Furthermore, note that the values of $T_{\rm eff}$, $\log g$, [M/H], and [α/M] will not be identical to their corresponding values in Ho et al. (2016) for several reasons: they are on the uncalibrated APOGEE label scale, and there have been various changes in our procedure (masking 50% of the spectrum, the inclusion of photometry, fitting for additional labels).

Table 1: Excerpt from the online table of stellar labels (T_{eff}, log g, [M/H], [C/M], [N/M], [α/M], A_k, mass, and age) for 454,364 stars, inferred by *The Cannon*. Column 1 is the LAMOST ID of the object, Columns 2-3 are the position of the object in RA and Dec, Columns 4-10 are the labels from *The Cannon*, Columns 11-12 are the estimated masses and ages, Columns 13-19 are the formal errors on the Cannon-inferred labels from the covariance matrix in the model fit, Columns 20-21 are the estimated uncertainties on mass and age, and Columns 22-23 are the S/N of the spectrum and the reduced χ^2 of the model fit. Note that the reduced χ^2 values are low by a factor of $\sim$3 because the random component of the errors in the LAMOST spectra is overestimated.

LAMOST ID	RA (deg)	Dec (deg)	T_{eff} (K)	log g (dex)	[M/H] (dex)	[C/M] (dex)	[N/M] (dex)	[α/M] (dex)	A_k mag	Mass ($M_\odot$)	log(Age) dex
spec-55859-F5902_sp01-034	331.92	-1.78	4794	3.22	-0.507	0.0645	-0.0242	0.228	0.0540	0.78	1.0
spec-55859-F5902_sp03-209	331.14	0.853	4620	2.88	-0.347	0.0984	0.107	0.220	0.0131	1.0	0.85
spec-55859-F5902_sp06-160	334.27	-0.159	4240	2.23	-0.293	0.0734	0.102	0.208	0.148	1.3	0.66
spec-55859-F5902_sp08-146	333.41	-0.397	4895	3.29	-0.337	-0.0221	-0.0243	0.212	0.0293	1.2	0.65

Table 2: Continued from Table 1

LAMOST ID	$\sigma(T_{eff})$ (K)	σ(log g) (dex)	σ([M/H]) (dex)	σ([C/M]) (dex)	σ([N/M]) (dex)	σ([α/M]) (dex)	$\sigma(A_k)$ (mag)	σ(Mass) ($M_\odot$)	σ(log(Age)) (dex)
spec-55859-F5902_sp01-034	3290	0.010	0.0027	0.0043	0.0089	0.00066	0.00036	0.33	0.34
spec-55859-F5902_sp03-209	73.8	0.00031	8.0e-5	0.00012	0.0003	2.67e-5	5.2e-5	0.097	0.12
spec-55859-F5902_sp06-160	65.0	0.0004	0.0001	7.9e-5	0.00014	4.93e-5	7.0e-5	0.43	0.34
spec-55859-F5902_sp08-146	5150	0.016	0.0047	0.0042	0.0058	0.0015	0.00041	0.47	0.45

Table 3: Continued from Table 2

LAMOST ID	SNR	Red. χ^2
spec-55859-F5902_sp01-034	33.7	0.44
spec-55859-F5902_sp03-209	169	1.7
spec-55859-F5902_sp06-160	130	1.2
spec-55859-F5902_sp08-146	19.9	0.51

3.1. Astrophysical Verification of Inferred Ages

We now investigate whether our inferred age values seem astrophysically plausible. Figure 5 shows the ([M/H], [α/M]) plane color-coded by age for $\sim 40,000$ objects with S/N > 80. Shown is the mean age in each bin, weighted by the estimated uncertainty in the age measurement, for bins with a minimum of 20 objects. We see an astrophysically sensible age gradient with changing abundances, from the young, low-[α/M] sequence to the old, high-[α/M] sequence. This is qualitatively very similar to the gradient seen from small high-resolution datasets of main sequence turn-off stars in the solar neighborhood (e.g. Haywood et al. (2013)).

Furthermore, as Figure 6 shows, our masses and ages (for the reference objects) are in remarkable agreement with the masses and ages from the Ness et al. (2016) catalog, determined via a rather different approach. In that approach (x axis) masses were measured directly from APOGEE spectra ($R \sim 22,500$) and ages were estimated via isochrone fitting. In our approach (y axis) we measured [C/M] and [N/M] directly from LAMOST spectra ($R \sim 1,800$) and used the relations in Martig et al. (2016) to estimate ages. This agreement supports the plausibility of our estimates.

Finally, Figure 7 shows the enhanced spatial distribution of our sample over that from 70,000 stars in APOGEE (Ness et al. 2016). As expected, younger stars are concentrated towards the disk mid-plane, and older stars extend to a larger scale height away from the disk and into the bulge and halo.

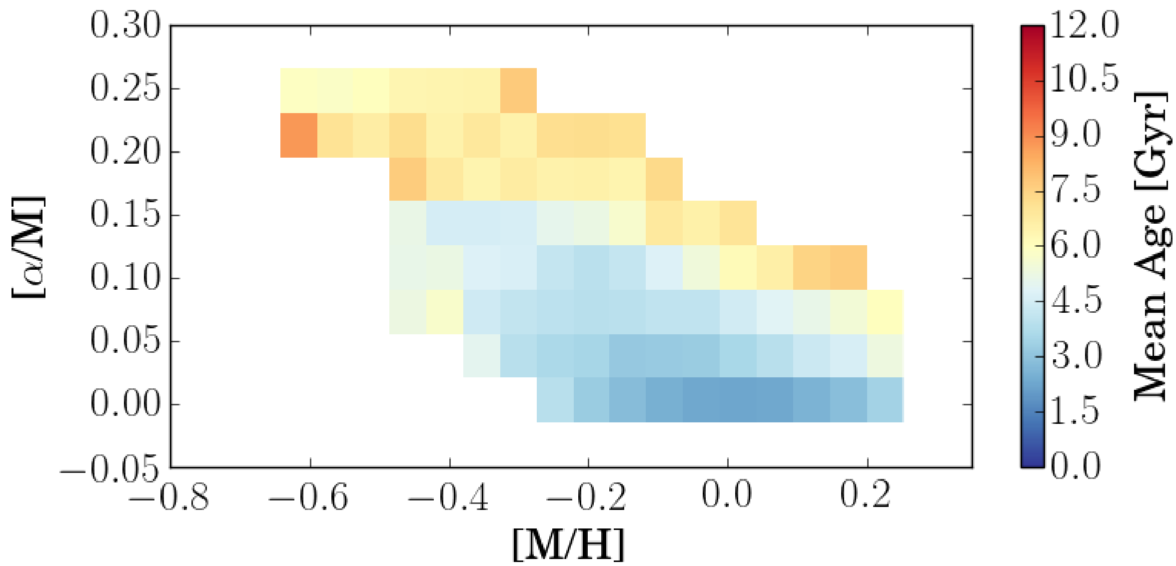

Fig. 5.— Cannon age estimates for LAMOST giants: the [α/M]-[M/H] plane color-coded by the mean age (weighted by the estimated age uncertainties) in each bin. We excised objects with S/N$<$ 80 (leaving 42420 objects) and only show bins with $>$ 20 objects. The gradient from young α-poor to old α-rich stars is astrophysically very plausible.

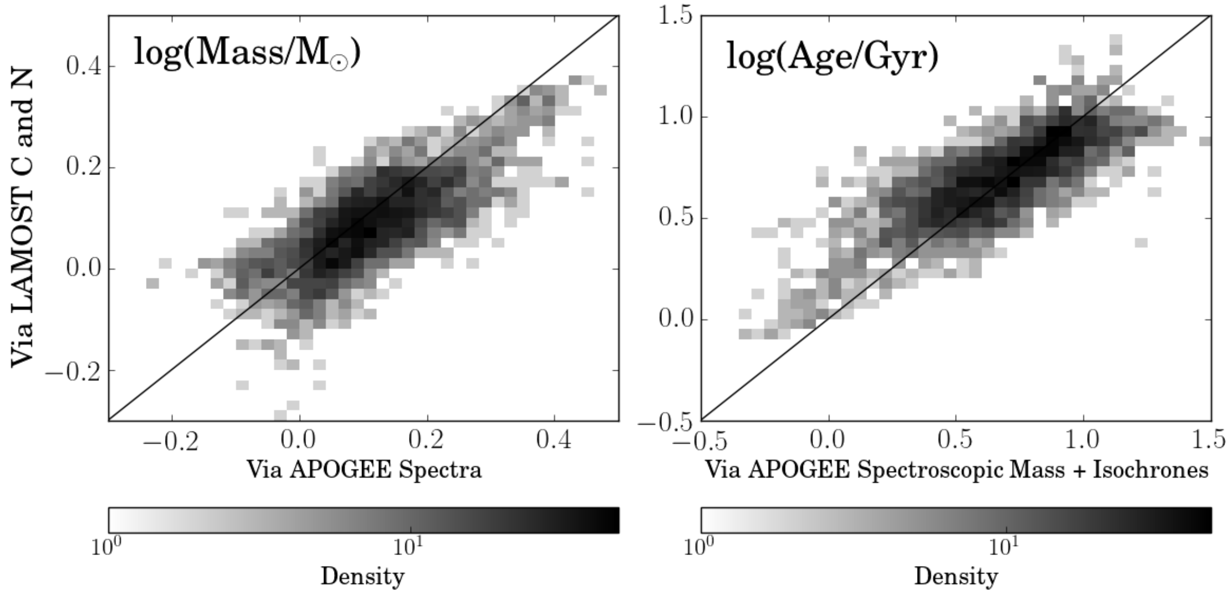

Fig. 6.— For the 6215 objects in common with the APOGEE mass and age catalog in Ness et al. (2016): comparison between our estimates (inferred via [C/M] and [N/M] abundances) and the Ness et al. (2016) mass and age estimates (inferred via spectroscopic mass measurements and isochrone fitting). The left panel shows the comparison for mass and the right panel shows the comparison for age. The agreement with the Ness et al. (2016) values despite the two very different approaches supports the plausibility of our measurements.

4. Discussion

Using a data-driven approach to spectral modeling, and fitting for all labels simultaneously, we find that we can measure accurate and precise carbon and nitrogen abundances from low-resolution ($R \sim 1800$) LAMOST spectra. For post dredge-up giants, as in the sample from Martig et al. (2016), these [C/M] and [N/M] measurements enable mass and age estimates across the sky, to 0.08 dex in mass and to 0.2 dex in age. With this new set of ages, we have a very different spatial sampling than APOGEE: we have essentially tied in-the-disk and off-the-disk ages onto the same scale, as LAMOST has a much better sampling of the thick disk than APOGEE.

The success of our data-driven approach in extracting information on [C/M] and [N/M] from blended regions (see Figure 4) holds promise for a natural extension of this work to measuring additional individual element abundances. *The Cannon* has already been successful at measuring individual abundances from APOGEE spectra, in part because the model is not restricted to unblended element windows (Ness et al. 2016; Casey et al. 2016;

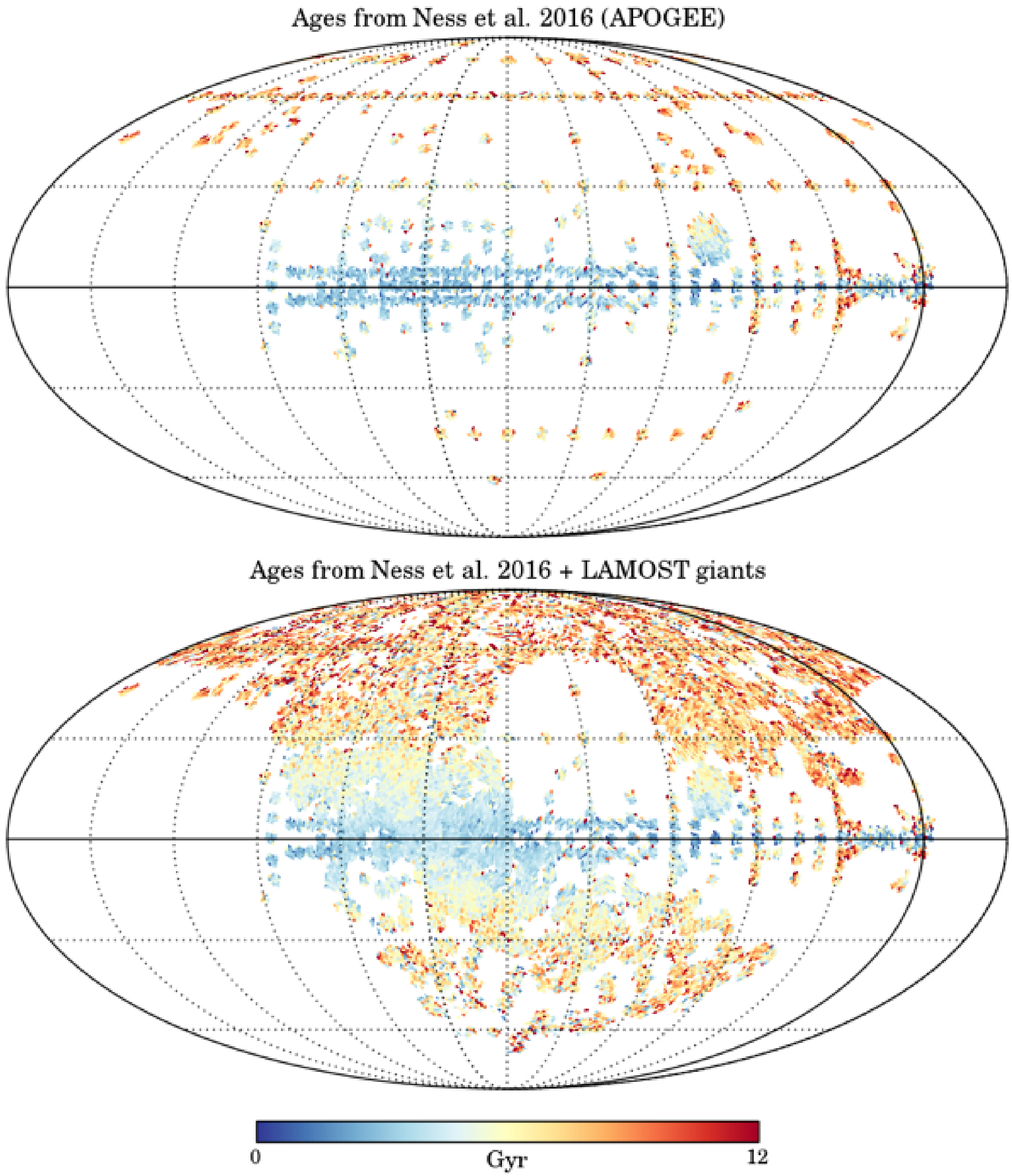

Fig. 7.— The distribution on the sky (in Galactic coordinates) of stars with age measurements: the top panel shows the sample from Ness et al. (2016) (∼70,000 objects) and the bottom panel overlays these values with 230,901 ages inferred via [C/M] and [N/M] by *The Cannon* from the LAMOST spectra. The much more extensive area coverage of the LAMOST data is immediately apparent.

Hogg et al. 2016). Indeed, Ting et al. 2016b predicted using theoretical models that spectra of comparable resolution to LAMOST should not only contain sufficient information to precisely constrain [C/M] and [N/M], but also a large suite of other individual element abundances, such as aluminum, calcium, manganese, and nickel.

For the purpose of [C/M] and [N/M] measurement in this work, it was helpful to apply broad masks to the spectra, to fully remove telluric and interstellar absorption features. Depending on which spectral regions encode information on [X/H], however, fitting for these additional labels would likely require more precise masking in order to avoid removing important signatures. The quality of the data reduction may limit which individual [X/H] can be returned.

Finally, at nearly identical values of $\{T_{\rm eff}, \log g, [Fe/H]\}$ on the giant branch, mass, or age, is highly predictive of luminosity. Age constraints to 0.2 dex could therefore be useful for improving estimates of stellar luminosity, and thus distance.

The code used to produce the results described in this paper was written in Python and is available online in an open-source repository.[2]

We are grateful to Evan Kirby (Caltech) for his careful reading and constructive feedback on the manuscript. It is furthermore a pleasure to thank Jon Bird (Vanderbilt), Sven Buder (MPIA), Andy Casey (IoA Cambridge), Christina Eilers (MPIA), Lynne Hillenbrand (Caltech), Sarah Martell (UNSW), and Marie Martig (MPIA) for helpful discussions and assistance. AYQH is grateful to the community at the MPIA for their support and hospitality during the period in which much of this work was performed.

AYQH was supported by a National Science Foundation Graduate Research Fellowship under Grant No. DGE1144469. MKN and HWR have received funding for this research from the European Research Council under the European Union's Seventh Framework Programme (FP 7) ERC Grant Agreement n. [321035]. DWH was partially supported by the NSF (grant IIS-1124794), NASA (grant NNX08AJ48G), and the Moore-Sloan Data Science Environment at NYU. CL acknowledges the Strategic Priority Research Program "The Emergence of Cosmological Structures" of the Chinese Academy of Sciences, Grant No. XDB09000000, the National Key Basic Research Program of China 2014CB845700, and the National Natural Science Foundation of China (NSFC) grants No. 11373032 and 11333003.

Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences.

[2]www.github.com/annayqho/TheCannon

Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatory of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

This research has made use of the APASS database, located at the AAVSO web site. Funding for APASS has been provided by the Robert Martin Ayers Sciences Fund.

This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

Facilities: Sloan(APOGEE spectrograph), LAMOST, WISE, AAVSO

REFERENCES

Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12

Allende Prieto, C., Beers, T. C., Wilhelm, R., et al. 2006, ApJ, 636, 804

Bedding, T. R., Huber, D., Stello, D., et al. 2010, ApJ, 713, L176

Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977

Bovy, J., Rix, H.-W., Schlafly, E. F., et al. 2015, arXiv:1509.05796

Casey, A. R., Hogg, D. W., Ness, M., et al. 2016, arXiv:1603.03040

Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, VizieR Online Data Catalog, 2246

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72

Epstein, C. R., Elsworth, Y. P., Johnson, J. A., et al. 2014, ApJ, 785, L28

García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2015, arXiv:1510.07635

Gilliland, R. L., Chaplin, W. J., Dunham, E. W., et al. 2011, ApJS, 197, 6

Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332

Haywood, M., Di Matteo, P., Lehnert, M. D., Katz, D., & Gómez, A. 2013, A&A, 560, A109

Henden, A., & Munari, U. 2014, Contributions of the Astronomical Observatory Skalnate Pleso, 43, 518

Henden, A. A., Templeton, M., Terrell, D., et al. 2016, VizieR Online Data Catalog, 2336

Ho, A. Y. Q., Ness, M. K., Hogg, D. W., et al. 2016, arXiv:1602.00303

Hogg, D. W., Casey, A. R., Ness, M., et al. 2016, arXiv:1601.05413

Holtzman, J. A., Shetrone, M., Johnson, J. A., et al. 2015, arXiv:1501.04110

Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, ApJ, 713, L79-L86

Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, A&A, 501, 1269

Kurucz, R. L. 1993, Kurucz CD-ROM, Cambridge, MA: Smithsonian Astrophysical Observatory, —c1993, December 4, 1993,

Kurucz, R. L., & Avrett, E. H. 1981, SAO Special Report, 391, 391

Kurucz, R. L. 1970, SAO Special Report, 309, 309

Kurucz, R. L. 1979, ApJS, 40, 1

Kurucz, R. L. 2005, Memorie della Societa Astronomica Italiana Supplementi, 8, 14

Liu, C., Deng, L.-C., Carlin, J. L., et al. 2014, ApJ, 790, 110

Liu, C., Fang, M., Wu, Y., et al. 2015, ApJ, 807, 4

Luo, A.L., Bai, Z. R., et al. 2015, RAA, in press

Majewski, S. R., Schiavon, R. P., Allende Prieto, C., et al. 2015, in preparation

Martell, S. L., Smith, G. H., & Briley, M. M. 2008, PASP, 120, 839

Martig, M., Fouesneau, M., Rix, H.-W., et al. 2015, arXiv:1511.08203

Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, ApJ, 808, 16

Ness, M., Hogg, D. W., Rix, H., et al. 2016, arXiv:1511.08204

Rix, H.-W., & Bovy, J. 2013, A&A Rev., 21, 61

Salaris, M., Pietrinferni, A., Piersimoni, A. M., & Cassisi, S. 2015, A&A, 583, A87

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Soderblom, D. R. 2010, ARA&A, 48, 581

Soderblom, D. R. 2015, Astrophysics and Space Science Proceedings, 39, 3

Stello, D., Huber, D., Bedding, T. R., et al. 2013, ApJ, 765, L41

Ting, Y.-S., Conroy, C., & Rix, H.-W. 2016, ApJ, 826, 83

Wan, J.-C., Liu, C., Deng, L.-C., et al. 2015, Research in Astronomy and Astrophysics, 15, 1166

Wilson, J. C., Hearty, F., Skrutskie, M. F., et al. 2010, Proc. SPIE, 7735, 77351C

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868-1881

Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011, Research in Astronomy and Astrophysics, 11, 924